

# Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 19, Number 13, November 2014

ISSN 1531-7714

---

## Analyzing Group Level Effects with Clustered Data Using Taylor Series Linearization

Francis L. Huang, *University of Missouri*

Clustered data (e.g., students within schools) are often analyzed in educational research where data are naturally nested. As a consequence, multilevel modeling (MLM) has commonly been used to study the contextual or group-level (e.g., school) effects on individual outcomes. The current study investigates the use of an alternative procedure to MLM: regression using Taylor series linearization (TSL) variance estimation. Despite the name, regressions using TSL are straightforward to conduct, can yield consistent and unbiased estimates and standard errors (given the appropriate conditions), and can be performed using a variety of commercially- and freely-available statistical software. I analyze a subsample of the High School and Beyond (HSB) dataset using MLM, regression using TSL, and ordinary least squares regression and compare results. In addition, 12,000 random samples are drawn from the HSB dataset of varying level-one and level-two sample sizes in order to compute biases in standard errors based on the different conditions. Sample R and SAS syntax showing how to run regressions using TSL are provided.

Multilevel modeling (MLM) has become a staple regression technique of choice for analyzing contextual effects using nested data within social science research. The use of multilevel models (also known as random coefficient models or more popularly known as hierarchical linear models) has grown in use over the years especially with educational research that often investigate the effects of schools or teachers on student outcomes. Though student-level (i.e., level-one units) outcomes are often evaluated, the treatment introduced in studies (e.g., a new curriculum to help raise math scores, a school-wide intervention to reduce bullying) is usually provided at the group level (i.e., level-two units) which requires researchers to properly account for the clustered nature of the data. MLM is a versatile and flexible analytic technique though not all nested data need to be analyzed using MLM (Huang, 2014).

Another older, readily-implemented but less well-known approach, at least in educational/psychological research circles, may effectively account for the clustered nature of the data when analyzing contextual effects: regression using Taylor series linearization (TSL; Rust, 1985). Despite the name, regressions using TSL are straightforward to perform, are available in both commercial (e.g., Mplus, SPSS) and free statistical software (e.g., R, SAS University), and may result in unbiased estimates and standard errors when the appropriate conditions are met.

This article compares results from models using a well-known dataset that investigates group-level effects as well as cross-level interactions using three regression techniques: standard ordinary least squares (OLS) regression, MLM, and regression using TSL. I compare regression results using a subset of the High School

and Beyond (HSB) dataset. In addition, 12,000 random samples drawn from the HSB dataset with differing level one and level two samples sizes are analyzed to investigate the results of the three techniques when sample size conditions are varied.

The objective of this article is to briefly introduce applied researchers to TSL as well as provide some guidance as to when regression using TSL may be appropriate in the analysis of clustered data. Alternative procedures for analyzing clustered data have received some attention (e.g., Arceneaux & Nickerson, 2009; Bliese, 2000; Harden, 2011), though studies have not specifically investigated TSL regression when analyzing group-level effects together with cross-level interactions (e.g., group-level variables interacting with individual-level variables). Cross-level interactions are particularly useful when investigating the effects of treatments administered at the group level which may vary in effect depending on student-level characteristics (e.g., “does the effect of the new curriculum [the ‘treatment’ introduced at the school or classroom level] on academic achievement differ based on a child’s socioeconomic status?”). The current study provides additional empirical evidence when TSL may be appropriate and provides example syntax which applied researchers may easily modify for their own use.

## Issues with Using OLS Regression in Analyzing Contextual Effects

Using OLS regression with nested data is problematic because observations within one group or cluster are more alike with each other compared to individuals in other groups or clusters, violating a well-known assumption of observation independence (Cohen, Cohen, West, & Aiken, 2003). For example, students within one school have more in common with each other compared to students in other schools potentially as a result of sharing the same teachers or school setting. Violating the assumption of observation independence has been known to lead to biased standard error estimates (i.e., estimates are too small or too large) which in turn can result in questionable inferences. When standard errors are underestimated, the probability of erroneously claiming statistical significance (i.e., a Type I error) increases (Clarke, 2008; Hox, 2002; Kreft & De Leeuw, 1998; Snijders & Bosker, 2012). The misestimation of standard errors is not a mere technical issue; if standard errors are consistently smaller than they should be (i.e., the

standard errors are biased downwards), traditional test statistics are artificially inflated which could result in the acceptance or rejection of important policy decisions based on study results.

In addition, when multilevel data are analyzed as a single-level dataset, the degrees of freedom of the level-two variables are artificially inflated as group-level variables are treated as repeating level-one variables in a model. For example, when analyzing data from 30 schools with 30 students per cluster (i.e., a total of  $n = 900$  students), school-level effects are evaluated using  $n - k$  (where  $k$  is the number of predictors) degrees of freedom even though in actuality there are only 30 schools so the degrees of freedom for level-one and level-two variables should be different. Again, greater degrees of freedom in this case may also contribute to increased Type I error rates.

Primarily, the violation of observation independence affects standard error estimates and not the regression coefficient (i.e., the  $b$ 's) estimates (Huang, 2014). Several studies using simulations (Harden, 2011; McNeish, 2014; Mundfrom & Schultz, 2001) and applied examples (Astin & Denson, 2009; Claessens, 2012; Newman, Newman, & Salzman, 2010) have repeatedly shown that estimates do not diverge much when either OLS or MLM is used. As a result, regression coefficient estimates using either method are generally unbiased and interest in the differences between methods usually focus on the proper estimation of the standard errors.

## Adjusting Standard Errors

Several standard error adjustment techniques have been developed and are known by various names that can correct standard errors depending on the type of regression assumption violated (e.g., heteroskedasticity, observation independence) (Petersen, 2008). Educational researchers should note that econometricians generally use a variety of adjusted standard error estimates that have names such as cluster adjusted, robust cluster adjusted, and bootstrap cluster adjusted standard errors (see Harden, 2011) among others which all differ from the standard errors resulting from standard OLS regression. In educational research, a commonly used manual method of adjusting standard errors when clustering is present is the design effect (DEFF) approach (Kish, 1965).

Several studies have shown how to manually adjust standard errors derived using OLS analyses of clustered

data (see Hahs-Vaughn, 2005; McCoach & Adelson, 2010). Even though DEFF adjustment may effectively adjust standard errors at the group level and reduce the likelihood of Type I errors for level-two coefficients, at the lowest level (i.e., level one), standard errors may be too conservative resulting in a high likelihood of Type II errors (Huang, 2014). As a result, researchers may have to deal with a trade-off where, depending on a combination of sample sizes at levels one and two, level-two estimates are unbiased while level-one estimates are too conservative. However, instead of having to manually adjust standard errors, an alternative procedure is readily available in several statistical programs, the use of the Taylor series linearization (TSL) variance estimation.

### Using Taylor Series Linearization

Taylor series linearization (TSL), also known as the Taylor approximation method, linearization, the delta method, or the propagation of variances, has been available for decades (Kish, 1965; Kish & Frankel, 1974; Laplante & Hébert, 2001; Rust, 1985) and is often used to analyze data collected through means other than simple random sampling (SRS). Complex sampling designs are typically used in nationally-representative surveys when data collected are not the product of a SRS. For practical purposes, nationally-representative surveys do not usually employ SRS and may select, for example, an initial set of clusters or groups (i.e., the primary sampling units) and from there, select observations within clusters using some other sampling procedure (e.g., stratified sampling, random sampling). Multiple levels of clusters may also be present such as when individuals from families are selected from neighborhoods within cities (Lumley, 2004). The National Center for Educational Statistics (NCES) routinely uses TSL when reporting results collected from its various national surveys (see Wang et al., 2011).

In the context of regression, coefficients estimates from models that use TSL are the same as those found when OLS regression is used except that standard errors will differ. TSL specifically makes use of the complex sampling design (e.g., the clusters, strata used) when making corrections to standard error estimates (Bell & McCaffrey, 2002; Lee & Forthofer, 2006). In the field of economics, sandwich variance estimators are often used to provide cluster robust standard errors which are similar to the standard error estimates

derived using TSL (Lumley, 2004). For applied researchers, the “mathematical explanation to this estimator [TSL] is a bit demanding” (Laplante & Hébert, 2001, p. 298) though fortunately is handled by the statistical software. TSL has been referred to as the gold standard for variance estimation in complex sample data (e.g., it is the default variance estimation procedure used in the SAS Survey procedures), it is relatively easy to implement, and has been used often in the regression-based analyses of large-sample data (Mukhopadhyay, An, Tobias, & Watts, 2008). Traditionally, the analysis of complex survey samples required specialized software (Lumley, 2004) such as SUDAAN (Research Triangle Institute, 2010). Nowadays however, various software packages can perform TSL even using freely-available programs (see Siller & Tompkins, 2006 for an overview and syntax for different commercial software packages).

### Performing a Regression Using TSL

For R users, researchers can install the Survey package (Lumley, 2014) which was developed for the analysis of complex survey samples that do not use SRS. For the analysis of simple clustered data (i.e., no stratification, no weights), researchers only need to specify the clustering variable using the `svydesign` function as well as the name of the dataset to be analyzed. After specifying the survey design, a regression can be performed using the `svyglm` function (e.g., math achievement is regressed on the mean socioeconomic status [SES] at the school level as well as the individual students SES). See Table 1 for example R code to analyze a downloadable version of the HSB data (available online at the indicated URL as a Stata .dta file, the `read.dta` function from the `foreign` package is used to import the dataset) which uses a school id (i.e., school) as the clustering variable. Beginning R users following the provided code should note that R is case sensitive.

Using R:

```
install.packages("survey")
library(survey)
install.packages("foreign")
library(foreign)
hsb<-
  read.dta(file="http://www.ats.ucla.edu/stat/paperexamples/singer/hsb12.dta")
design<-svydesign(id=~school,data=hsb)
tsl<-svyglm(mathach~meanses+ses,design)
summary(tsl)
```

By default, SAS uses TSL when the PROC SURVEY procedures are used and only requires the user to indicate what the clustering variable is with the cluster statement. Using SAS (after downloading the hsb12.dta into the c:\datam\ directory), SAS users can enter:

```
proc import
    datafile="c:\datam\hsb12.dta"
    out=hsb; run;

proc surveyreg data=hsb;
    cluster school;
    model mathach=meanses ses;
run;
```

To build cross-level interactions (i.e., level-two variable x level-one variable), modify the statement in R to read:

```
tsl<-svyglm(mathach~meanses+ses+
meanses:ses,design)
```

(the added meanses:ses term indicates a cross-level interaction between the average SES at the school level x the student's SES). In SAS, modify the model statement to read: model mathach=meanses ses meanses\*ses. Both programs allow for the automatic creation of interaction terms.

## Method

The purpose of the current study is to illustrate the differences and similarities using OLS regression, MLM, and regression using Taylor series linearization (TSL) variance estimation. I analyze a subsample of the 1982 High School and Beyond (HSB) dataset used by both Raudenbush and Bryk (2002) and Singer (1998).

The data are comprised of information from 7,185 students from 160 schools. The level-one outcome is the student's math achievement (MATHACH;  $M = 12.75$ ,  $SD = 6.88$ ) score and included as a covariate is a student's SES ( $M = 0.00$ ,  $SD = 0.78$ ). At level two (i.e., the school level), the aggregated SES ( $M = -0.00$ ,  $SD = 0.41$ ) is included as well as a SECTOR variable dummy coded as a 1 ( $n = 70$ ; Catholic school) and a 0 ( $n = 90$ ; non-Catholic school). The intraclass correlation, which indicates the amount of variability attributable to the group level, was .18 (see Raudenbush & Bryk, 2002, pp. 68-69, for a more detailed description of the sample).

Following the final full model specifications of Singer (1998, p. 339), the following equations are modeled using MLM:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij} - \overline{SES}_j) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}MEANSES_j + \gamma_{02}SECTOR + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}MEANSES_j + \gamma_{12}SECTOR$$

where  $Y_{ij}$  represents the math achievement score of student  $i$  in school  $j$ . The variable indicates that the students' SES is group-mean centered where the school's average SES is subtracted from the student's SES. The combined level-one and level-two equations result in:

$$Y_{ij} = \gamma_{00} + \gamma_{01}MEANSES_j + \gamma_{02}SECTOR_j$$

$$+ \gamma_{10}(SES_{ij} - \overline{SES}_j)$$

$$+ \gamma_{11}MEANSES_j(SES_{ij} - \overline{SES}_j)$$

$$+ \gamma_{12}SECTOR_j(SES_{ij} - \overline{SES}_j) + u_{0j} + r_{ij}$$

Both  $\gamma_{01}$  and  $\gamma_{02}$  represent the association of the level-two effects with math achievement while controlling for all other variables in the model. The  $\gamma_{10}$  term models the effect of the level-one predictor. Both  $\gamma_{11}$  and  $\gamma_{12}$  model the effects of the cross-level interactions. The OLS and TSL regressions, for purposes of comparability, have a similar equation except that the random component,  $u_{0j}$ , is not included.

Analyses were done using both R and SAS. The nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2014) for R was used to run MLM using restricted maximum likelihood. The survey package (Lumley, 2014) was used to run the regressions using TSL variance estimation by specifying the school as the cluster variable. The first set of analyses compare the regression coefficients and standard error estimates with each other.

However, as applied researchers collecting their own data may not always have access to a large sample, I investigate differences in regression techniques using a smaller sample. Using R, I conducted a bootstrap simulation to draw repeated samples from the HSB dataset while manipulating sample sizes at both level one ( $n = 5, 10, 15, 20$ ) and level two ( $j = 10, 30, 50$ ) to provide a total of 12 conditions. Similar to other simulation studies, 1,000 samples were drawn for each condition (Clarke, 2008; Hox & Maas, 2005) resulting in the 12,000 datasets (i.e., 4 level-one conditions x 3 level-two conditions x 1,000 replications). Level-two sample sizes were selected based on research that investigated sample size requirements at the group level (Hox & Maas, 2005; Kreft & De Leeuw, 1998; Snijders & Bosker, 2012).

As population parameters estimated using OLS regression, regression using TSL, and MLM are all relatively unbiased as shown by numerous studies (Astin & Denson, 2009; Harden, 2011; Huang, 2014; McNeish, 2014; Mundfrom & Schultz, 2001; Newman et al., 2010), the current study focused on the standard error estimates which is of primary concern when studying contextual effects. To evaluate standard error performance, I computed the relative bias of the standard error estimates compared to the empirical standard errors which is the standard deviation of the parameter estimates across the 1,000 replication for a given condition. Bias is computed as the difference of the estimated standard error less the empirical standard error divided by the empirical standard error multiplied by 100 (i.e.,  $\text{bias} = \frac{\hat{\theta} - \theta}{\theta} \times 100$ ). Based on prior Monte Carlo simulation studies, bias in standard errors less than  $\pm 10\%$  were not considered problematic (Muthén & Muthén, 2002).

## Results

Parameter estimates, modeled with R, using OLS regression, regression using TSL variance estimation, and MLM all yielded comparable estimates for level-one, level-two, and cross-level interactions terms (see Table 1). Standard errors for level-one estimates as well as the two cross-level interactions were similar for all models. However, standard errors for OLS regression coefficients were much smaller than either estimates using MLM or TSL but only for the level-two standard errors. Results were similar when done using PROC MIXED (for the MLM model) and PROC SURVEYREG (for regression using TSL) using SAS.

**Bootstrapped Standard Errors.** Even though model results using the entire sample showed that regressions using TSL variance estimation and MLM yielded comparable estimates and standard errors, I investigated if the results were similar when level-one and level-two sample size conditions are varied. Tables 2 to 4 show the bias in the standard error estimates across the different conditions. All MLM models converged (i.e., allowed admissible solutions) with the exception of five models (out of 1,000) for the smallest level-2 sample size condition ( $j = 10$ ).

Table 1. Comparison of Parameter Estimates and Standard Errors Using Multilevel Modeling (MLM), Ordinary Least Squares Regression (OLS), and Regression Using Taylor Series Linearization (TSL) Variance Estimation

| Variable                         | MLM   |      | TSL   |      | OLS   |      |
|----------------------------------|-------|------|-------|------|-------|------|
|                                  | Est   | SE   | Est   | SE   | Est   | SE   |
| Intercept                        | 12.11 | 0.20 | 12.10 | 0.17 | 12.10 | 0.11 |
| Mean SES ( $\gamma_{01}$ )       | 5.34  | 0.37 | 5.17  | 0.34 | 5.17  | 0.19 |
| Sector ( $\gamma_{02}$ )         | 1.21  | 0.31 | 1.27  | 0.30 | 1.27  | 0.16 |
| SES ( $\gamma_{10}$ )            | 2.94  | 0.15 | 2.94  | 0.15 | 2.94  | 0.15 |
| Mean SES x SES ( $\gamma_{11}$ ) | 1.04  | 0.29 | 1.04  | 0.33 | 1.04  | 0.30 |
| Sector x SES ( $\gamma_{12}$ )   | -1.64 | 0.23 | -1.64 | 0.24 | -1.64 | 0.24 |

Notes. SES = socioeconomic status. SES at the student level is group-mean centered. All estimates are statistically significant (all  $p$ 's < .001).

Table 2. Bias in Level Two Standard Errors Using Ordinary Least Squares (OLS) Regression, Regression Using Taylor Series Linearization (TSL) Variance Estimation, and Multilevel Modeling (MLM) by Level One and Level Two Sample Size Conditions

| Lvl | Lvl | $\gamma_{01}$ |       |      | $\gamma_{02}$ |       |      |
|-----|-----|---------------|-------|------|---------------|-------|------|
|     |     | OLS           | TSL   | MLM  | OLS           | TSL   | MLM  |
| 10  | 5   | -12.8         | -28.4 | 1.4  | -19.5         | -29.6 | -6.3 |
|     | 10  | -17.5         | -23.8 | 4.9  | -20.4         | -20.1 | 1.3  |
|     | 15  | -26.5         | -25.8 | 1.1  | -28.6         | -22.4 | -2.2 |
|     | 20  | -35.5         | -31.1 | -8.0 | -36.2         | -27.9 | -9.3 |
| 30  | 5   | -4.5          | -7.0  | 7.6  | -7.6          | -2.6  | 4.1  |
|     | 10  | -14.8         | -7.3  | 6.0  | -16.9         | -1.0  | 3.4  |
|     | 15  | -17.7         | -1.7  | 12.6 | -24.5         | -0.4  | 3.2  |
|     | 20  | -25.1         | -2.5  | 10.5 | -27.5         | 3.2   | 6.1  |
| 50  | 5   | -0.6          | 1.0   | 11.7 | -7.6          | 0.6   | 3.9  |
|     | 10  | -8.7          | 3.4   | 14.5 | -14.8         | 4.9   | 6.9  |
|     | 15  | -12.1         | 7.8   | 19.8 | -18.2         | 10.2  | 11.4 |
|     | 20  | -18.1         | 8.2   | 20.7 | -24.4         | 9.7   | 10.5 |

Notes. Shaded numbers are underestimated by over 10%. Underlined numbers are overestimated by over 10%.

**Level-two Standard Errors.** Table 2 shows the standard error bias for the level-two (i.e., school level) coefficients (i.e.,  $\gamma_{01}$ ,  $\gamma_{02}$ ). As could be expected based on prior research, OLS standard errors for the group-level variables were mostly underestimated (i.e., too small) and biased downwards. However, regression using TSL were also biased downward for the smallest level-two sample size condition ( $j = 10$ ). In general, the standard errors for the level-two variables for the

regressions using TSL and MLM were comparable when group level sample size was at least 30. MLM standard errors for level-two variables though were slightly more conservative.

Table 3. Bias in Level One Standard Errors Using Ordinary Least Squares (OLS) Regression, Regression Using Taylor Series Linearization (TSL) Variance Estimation, and Multilevel Modeling (MLM) by Level One and Level Two Sample Size Conditions

| Sample size |         | $\gamma_{11}$ |       |             |
|-------------|---------|---------------|-------|-------------|
| Level 2     | Level 1 | OLS           | TSL   | MLM         |
| 10          | 5       | -6.0          | -28.9 | -8.4        |
|             | 10      | -0.3          | -24.8 | -1.5        |
|             | 15      | 0.7           | -22.6 | -1.5        |
|             | 20      | -0.3          | -22.1 | -1.6        |
| 30          | 5       | 0.0           | -7.9  | -1.1        |
|             | 10      | 5.6           | -2.7  | 3.6         |
|             | 15      | 3.9           | -4.7  | 1.7         |
|             | 20      | 8.9           | 0.3   | 7.2         |
| 50          | 5       | -0.4          | -4.0  | -1.3        |
|             | 10      | 2.5           | -2.0  | 1.2         |
|             | 15      | 6.9           | 1.2   | 4.4         |
|             | 20      | <u>13.9</u>   | 8.3   | <u>11.3</u> |

Notes. Shaded numbers are underestimated by over 10%. Underlined numbers are overestimated by over 10%.

**Level-one Standard Errors.** Standard errors for the level-one coefficient was generally unbiased for all conditions and type of regressions used, with the exception of regression using TSL when the number of groups was small ( $j = 10$ ). Standard errors for regressions for  $\gamma_{11}$  using TSL variance estimation were consistently underestimated by 22-29%.

**Cross-level Interaction Standard Errors.** Similar to the level-one standard errors, the standard errors for the regressions using TSL were also consistently underestimated for the smallest group-level sample size condition ( $j = 10$ ). However, even when  $j = 30$ , when level-one sample sizes were small (i.e.,  $n = 5$ ), TSL still underestimated the standard errors. On the other hand, the standard errors when OLS regression was used were relatively unbiased and comparable to the MLM standard errors. In other words, if cross-level interactions are of primary interest, using standard OLS regression, regardless of level-one or level-two sample size, did not result in biased standard errors.

Table 4. Bias in Cross-level Interaction Standard Errors Using Ordinary Least Squares (OLS) Regression, Regression Using Taylor Series Linearization (TSL) Variance Estimation, and Multilevel Modeling (MLM) by Level One and Level Two Sample Size Conditions

| Sample size |       | $\gamma_{11}$ |       |             | $\gamma_{12}$ |       |      |
|-------------|-------|---------------|-------|-------------|---------------|-------|------|
| Lvl 2       | Lvl 1 | OLS           | TSL   | MLM         | OLS           | TSL   | MLM  |
| 10          | 5     | -3.9          | -29.7 | -6.2        | -5.4          | -31.0 | -8.1 |
|             | 10    | -1.5          | -28.9 | -3.6        | 0.7           | -26.3 | -0.2 |
|             | 15    | -1.7          | -27.6 | -3.9        | -0.9          | -27.1 | -1.8 |
| 20          | 5     | -5.7          | -28.9 | -7.6        | 4.4           | -24.1 | 2.3  |
|             | 10    | -0.6          | -12.0 | -1.7        | 0.7           | -9.0  | -0.9 |
|             | 15    | 0.9           | -8.4  | -1.0        | 3.4           | -6.6  | 1.6  |
| 30          | 5     | 0.8           | -8.0  | -1.5        | 0.7           | -9.1  | -0.2 |
|             | 10    | 2.1           | -4.5  | -0.2        | 8.2           | -2.0  | 7.3  |
|             | 15    | 2.0           | -3.1  | 1.3         | 1.9           | -3.8  | 1.5  |
| 50          | 5     | -0.1          | -7.5  | -1.0        | 2.6           | -3.7  | 1.9  |
|             | 10    | 2.0           | -3.1  | 1.3         | 1.9           | -3.8  | 1.5  |
|             | 15    | 3.7           | 0.2   | 2.7         | 8.6           | 1.5   | 8.3  |
| 20          | 5.0   | 3.9           | 3.1   | <u>10.6</u> | 4.4           | 9.5   |      |

Notes. Shaded numbers are underestimated by over 10%. Underlined numbers are overestimated by over 10%.

## Discussion

Regression model results, derived through analyses of a subset of the HSB dataset, indicated that multilevel modeling (MLM) and regression using Taylor series linearization (TSL) variance estimation yielded similar parameter estimates and standard errors, regardless of level of analysis (i.e., level-one, level-two, cross-level interaction terms). However, level-two standard errors were much lower for the OLS regression results compared to either model using TSL or MLM, in line with prior research (Harden, 2011; McNeish, 2014). In general, when analyzing large datasets and the researcher is interested in level-two variables as well as cross-level interactions, either MLM or TSL will yield consistent (i.e., unbiased) estimates and standard errors. While MLM is a versatile and flexible technique, TSL regression may offer a simpler alternative, one which is straightforward to run (as shown by the provided syntax). As stated by Bell and McCaffrey (2002, p. 168), “many analysts [may] prefer the simplicity of standard regression models when random effects are not of direct interest.”

If smaller samples though are to be studied, researchers are cautioned with using TSL variance estimation when group-level sample sizes are small ( $j < 30$ ).

While TSL methods are widely used in the analysis of large, nested datasets (Mukhopadhyay et al., 2008), researchers gathering their own data may not have access to hundreds of schools. Our findings are consistent with Bell and McCaffrey's (2002) earlier study which showed that TSL standard errors may be severely biased downwards with small group sizes. Even with larger group sizes (i.e.,  $j = 30$ ), researchers should strive to have at least 10 subjects per group to avoid negatively biased standard errors when examining cross-level interactions (see Table 4). Even though regression using TSL is a viable alternative to MLM, larger group-level sample sizes may be needed.

Of note as well is that while level-two standard errors are underestimated using OLS regression, the cross-level interactions were consistent and unbiased. In some instances, standard errors using OLS regressions were actually larger (i.e., more conservative) compared to those derived using MLM (see Table 4). This was shown using the 12,000 samples drawn as well as when the main dataset used (see Table 1). While the use of OLS regression may be frowned upon when nested datasets are analyzed, the issue of Type I error has generally been shown to be associated with the level-two (i.e., group-level) coefficients and analysts should not conclude that all standard errors using OLS regression are underestimated (see Harden, 2011; Huang, 2014).

There are some limitation to consider in assessing the current study's results. First, as a simulation, the findings may be specific to the conditions of the data investigated. However, in modeling results, the current study did consider the effects of level-one, level-two, and cross-level interactions terms all together using a readily available dataset. Second, in order to be consistent with prior studies that have used the HSB dataset (e.g., Raudenbush & Bryk, 2002; Singer, 1998) in the teaching of MLM, I followed the exact same centering strategies used for the variables though the type of centering used may change results, which could be an area for future study. Raudenbush and Bryk (2002) as well as Enders and Tofighi (2007) though indicate that group-mean centering variables may be most appropriate when the level-one variables are of primary interest.

## Conclusion

The analysis of clustered data is commonly performed in the social and behavioral sciences and I

present an older, well-accepted, though not often used technique (in educational research circles) that is suitable for the analysis of clustered data: regression using Taylor series linearization (TSL). If large, nested datasets are investigated, such as those provided by the NCES, then using TSL or MLM should result in comparable and acceptable estimates when studying contextual or level-two effects. Taylor series linearization is often used in the analysis of nested datasets and is one of the options, in addition to using MLM, suggested by the NCES in their training sessions for the analysis of nationally-representative datasets.

However, with smaller datasets, TSL may be a good alternative to MLM when group size is at least 30 and there are 10 or more observations within each cluster. The current study has shown, that given the appropriate conditions described, regression using TSL variance estimation may result in unbiased estimates and TSL regression can be readily performed using a variety of commercial and free statistical software.

## References

- Arceneaux, K., & Nickerson, D. W. (2009). Modeling certainty with clustered data: A comparison of methods. *Political Analysis*, 17, 177–190. doi:10.1093/pan/mpp004
- Astin, A. W., & Denson, N. (2009). Multi-campus studies of college impact: which statistical method is appropriate? *Research in Higher Education*, 50, 354–367. doi:10.1007/s11162-009-9121-3
- Bell, R., & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28, 169–182.
- Bliese, P. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. Bollen & J. Long (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.
- Claessens, A. (2012). Kindergarten child care experiences and child achievement and socioemotional skills. *Early Childhood Research Quarterly*, 27, 365–375. doi:10.1016/j.ecresq.2011.12.005
- Clarke, P. (2008). When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology and Community Health*, 62, 752–758. doi:10.1136/jech.2007.060798
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge.

- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121–138. doi:10.1037/1082-989X.12.2.121
- Hahs-Vaughn, D. L. (2005). A primer for using and understanding weights with national datasets. *The Journal of Experimental Education, 73*, 221–248. doi:10.3200/JEXE.73.3.221-248
- Harden, J. J. (2011). A bootstrap method for conducting statistical inference with clustered data. *State Politics & Policy Quarterly, 11*, 223–246. doi:10.1177/1532440011406233
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Hox, J. J., & Maas, C. J. M. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86–92.
- Huang, F. L. (2014). Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education*. Advance online publication. doi:10.1080/00220973.2014.952397
- Kish, L. (1965). *Survey sampling*. New York, NY: Wiley.
- Kish, L., & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society. Series B (Methodological), 36*, 1–37.
- Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage.
- Laplante, B., & Hébert, B.-P. (2001). An introduction to the use of linear models with correlated data. *Canadian Studies in Population, 28*, 287 – 311.
- Lee, E. S., & Forthofer, R. N. (2006). *Analyzing Complex Survey Data*. Thousand Oaks, CA: Sage.
- Lumley, T. (2014). *Survey: Analysis of complex survey samples. R Package version 3.30*.
- McCoach, D. B., & Adelson, J. L. (2010). Dealing with dependence (part I): Understanding the effects of clustered data. *Gifted Child Quarterly, 54*, 152–155. doi:10.1177/0016986210363076
- McNeish, D. M. (2014). Analyzing clustered data with OLS regression: The effect of a hierarchical data structure. *Multiple Linear Regression Viewpoints, 40*, 11–16.
- Mukhopadhyay, P., An, A., Tobias, R., & Watts, D. (2008). *Try, try again: Replication-based variance estimation methods for survey data analysis in SAS® 9.2*. Paper presented at the SAS Global Forum 2008.
- Mundfrom, D. J., & Schultz, M. R. (2001). A comparison between hierarchical linear modeling and multiple linear regression in selected data sets. *Multiple Linear Regression Viewpoints, 27*, 3–11.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 599–620. doi:10.1207/S15328007SEM0904\_8
- Newman, D., Newman, I., & Salzman, J. (2010). Comparing OLS and HLM models and the questions they answer: Potential concerns for type VI errors. *Multiple Linear Regression Viewpoints, 36*, 1–10.
- Petersen, M. A. (2008). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies, 22*(1), 435–480. doi:10.1093/rfs/hhn053
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2014). *nlme: Linear and Nonlinear Mixed Effects Models*. Retrieved from <http://CRAN.R-project.org/package=nlme>
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Research Triangle Institute. (2010). SUDAAN language manual release 10.0. Research Triangle Park, NC: Research Triangle Institute.
- Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics, 1*, 381–397.
- Siller, A. B., & Tompkins, L. (2006, March). The big four: Analyzing complex sample survey data using SAS®, SPSS®, STATA®, and SUDAAN®. In Proceedings of the Thirty-first Annual SAS® Users Group International Conference (pp. 26-29).
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*, 323–355. doi:10.2307/1165280
- Snijders, T., & Bosker, R. (2012). *Multilevel analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- Wang, X., Henning, A., Cui, W., Huang, F., Armstrong, S., Kang, K., Boyer, J., & Robers, S. (2011). *NCES handbook of survey methods (NCES 2011-609)*. In S. Burns, X. Wang, & A. Henning (Eds.). Washington, D.C.: U.S. Department of Education, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubs2011/2011609.pdf>



**Citation:**

Huang, Francis L. (2014). Analyzing Group Level Effects with Clustered Data Using Taylor Series Linearization. *Practical Assessment, Research & Evaluation*, 19(13). Available online: <http://pareonline.net/getvn.asp?v=19&n=13>

**Corresponding Author:**

Francis L. Huang  
Educational, School, and Counseling Psychology  
University of Missouri  
16 Hill Hall  
Columbia, MO 65211  
Huangf [at] missouri.edu